DOCUMENT RESUME

ED 289 503 IR 012 970

AUTHOR Ellington, Henry
TITLE Student Assessment. Teaching and Learning in Higher Education, 20.
INSTITUTION Scottish Central Institutions Committee for Educational Development.
SPONS AGENCY Robert Gordon's Inst. of Technology, Aberdeen (Scotland).
PUB DATE 87
NOTE 21p.; For related guides, see IR 012 971-74.
PUB TYPE Guides - Non-Classroom Use (055) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Criterion Referenced Tests; Foreign Countries; Higher Education; Norm Referenced Tests; *Student Evaluation; *Test Construction; *Test Format

ABSTRACT

Intended to provide a broad introduction to the subject of student assessment, this booklet begins by discussing the role of assessment in any systematic approach to course or curriculum design and explains the difference between assessment and evaluation. Four basic features of a good student assessment procedure are then discussed, i.e., validity, reliability, practicability, and fairness/usefulness. The differences between criterion-referenced assessment and norm-referenced assessment are also explained, and guidelines for constructing a test or other form of assessment are presented. The booklet concludes with discussions of five methods commonly used to carry out student assessment in terms of their design characteristics, functions, and strengths and weaknesses: (1) traditional extended writing tests; (2) objective tests; (3) practical tests; (4) unobtrusive assessment; and (5) self and peer assessment. An annotated list of three items recommended for further reading is included. (MES)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Student Assessment

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Henry Ellington

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This booklet was first published internally in Robert Gordon's Institute of Technology, Aberdeen as part of the Institute's staff development programme.

The booklet was produced by Dr Henry Ellington of RGIT's Educational Technology Unit, who adapted it from material originally published in "A Handbook of Educational Technology", by Fred Percival and Henry Ellington (Kogan Page, London; 1984).

CICED gratefully acknowledges the co-operation of RGIT, Kogan Page and the authors in the publication of the present edition of the booklet".

© CICED 1987

Student Assessment

Introduction

This booklet provides a broad introduction to the subject of *student assessment*. It begins by discussing the role of student assessment in any systematic approach to course or curriculum design, explaining the difference between *assessment* and *evaluation* (the subject of another booklet in the series). Next, it discusses the basic characteristics that any worthwhile student assessment scheme should possess and explains the difference between the two main approaches to student assessment – *criterion-referenced assessment* and *norm-referenced assessment*. Next, it offers guidance on how to set about constructing a test or other form of assessment. It then examines the various methods that are commonly used to carry out student assessment, discussing them in terms of their design characteristics, their functions and their respective strengths and weaknesses. Finally, mention is made of the potential of self and peer assessment.

Three of the most important assessment methods discussed are examined in much greater detail in three separate booklets that form a sequel to the present booklet – "Multiple-choice questions", "Short-answer questions" and "Essay-type questions".

The role of assessment in an instructional system

In the booklet on 'Educational objectives', it was shown that the process of course or curriculum development can be represented schematically by Figure 1.

As can be seen, the process in basically cyclic in nature, with the first three stages being:

(i) the formulation of a clear set of *objectives* for the course or curriculum;

(ii) the selection of appropriate *instructional methods* for achieving these objectives within the context of the course or curriculum;

(iii) the *implementation* of the course or curriculum.

Detailed guidance on how to carry out stages (i), (ii) and (iii) was given in other booklets in this series, and, in the present booklet, we will start to examine the fourth and final stage of the course or curriculum development process – the *assessment and evaluation* stage.

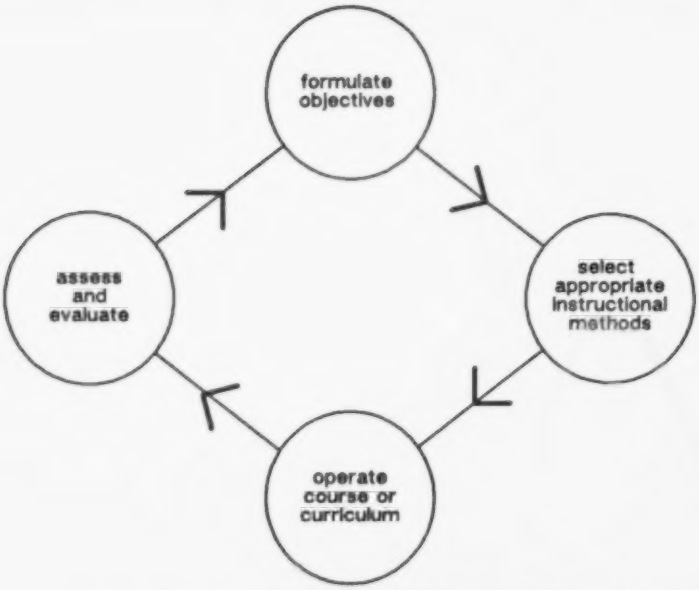

Figure 1: schematic representation of the systems approach to course or curriculum development.

The difference between assessment and evaluation

At this point, it would probably be useful to explain exactly what we mean by the terms *assessment* and *evaluation*. Although the two terms are often considered to be virtually synonymous when used in common parlance, they have radically different connotations when used in an educational or training context.

By *assessment*, first of all, we mean those activities that are designed to measure learner achievement brought about as a result of an instructional programme of some sort.

Evaluation, on the other hand, refers to a series of activities that are designed to measure the effectiveness of the instructional system as a whole.

Clearly, the two processes are fairly closely related, since the results of student *assessment* constitute one of the most important sets of data that should be taken into account in the *evaluation* of any course or curriculum. Both are also closely related to the *objectives* of the course or curriculum, since they are both basically concerned with determining the extent to which these objectives have (or have not) been achieved. Indeed, one cogent argument for articulating the objectives of a course or curriculum in fairly detailed (preferably

behavioural) form whenever possible is that this is generally of considerable assistance both in assessing the students and in evaluating the course or curriculum, since the designer should (as a result of writing the objectives in this way) have a fairly clear idea of the behaviour that is to be measured. Conversely, the feedback obtained from the results of properly designed assessment and evaluation procedures often demonstrates a need for changes in the actual objectives of the course or curriculum, as well as in the methods adopted for trying to achieve these.

Desirable characteristics of student assessment procedures

We will now turn our attention to the basic features that should characterise a 'good' student assessment procedure. Such a procedure should, ideally, be *valid*, *reliable*, *practicable*, and *fair and useful to students*. Let us now discuss these in turn.

Validity

A *valid* assessment procedure is one which actually tests what it sets out to test, i.e. one which accurately measures the behaviour described by the objective(s) under scrutiny. Obviously, no-one would *deliberately* construct an assessment item to test trivia or irrelevant material, but it is surprising just how often non-valid test items are in fact used – e.g. questions that are intended to test recall of factual material but which actually test the candidate's powers of reasoning, or questions which assume a level of pre-knowledge that the candidates do not possess.

As we will see later in the review of assessment methods, validity-related problems are a common weakness of many of the more widely-used methods. For example, a simple science question given to 14-year old schoolchildren ('Name the products of the combustion of carbon in an adequate supply of oxygen') produced a much higher number of correct answers when the word 'combustion' was replaced by 'burning'. This showed that the original question had problems of validity in that it was, to some extent, testing language and vocabulary skills rather than the basic science involved.

Reliability

The *reliability* of an assessment procedure is a measure of the consistency with which the question, test or examination produces the same results under different but comparable conditions. A reliable assessment item gives reproducible scores with similar populations of students, and is therefore as independent of the

characteristics and vagaries of individual markers as possible. This is often difficult to achieve in practice.

It is obviously important to have reasonably reliable assessment procedures when a large number of individual markers assess the same question (e.g. in national school examinations). A student answer which receives a score of 75 per cent from one marker and 35 per cent from another, for example, reveals a patently unreliable assessment procedure.

To help produce reliability, the questions which comprise a student assessment should (ideally) test only one thing at a time and give the candidates no choice. The assessment should also adequately reflect the objectives of the teaching unit. Note that the reliability and validity factors in an assessment are in no way directly linked – a test or examination, for example, may be totally reliable and yet have very low validity, and vice versa.

Practicability

For most purposes, assessment procedures should be realistically *practical* in terms of their cost, time taken, and ease of application. For example, with a large class of technicians being trained in electrical circuitry, it may only be convenient to use a paper-and-pencil test rather that set up numerous practical testing situations. It should be noted, however, that such compromises can, in some cases, greatly reduce the validity of the assessment.

Fairness and Usefulness

To be fair to all students, an assessment must accurately reflect the range of expected behaviours as described by the course objectives. It is also highly desirable that students should know exactly *how* they are to be assessed.

Indeed, it could be argued that students have a *right* to information such as the nature of the materials on which they are to be examined (i.e. content and objectives), the form and structure of the examination, the length of the examination, and the value (in terms of marks) of each component of the course.

Also, students should (ideally!) find assessments useful. Feedback from assessment can give a student a much better indication of his or her current strengths and weaknesses than he/she might otherwise have. In this respect, the non-return of assessment work to students greatly reduces its utility.

Criterion-referenced and norm-referenced assessment

Let us now turn our attention to the two basic (and contrasting) approaches that can be adopted to student assessment – *criterion-referenced assessment* and *norm-referenced assessment*.

Criterion-referenced assessment

Criterion-referenced assessment involves testing students in order to measure their performance in tasks described by a particular objective or set of objectives (the *criterion*). In any systems approach to education or training (which is invariably geared towards the achievement of clearly-specified objectives), it is normal to use some kind of criterion-referenced test for student assessment. In such a test, the relative performances of the various individuals in the class is of little consequence – indeed, in the unlikely event of the whole class demonstrating complete mastery of the objectives, this would simply indicate that a highly-successful teaching/learning system had been developed.

A good example of a criterion-referenced test is the standard driving test, in which the learner driver has to demonstrate a certain level of competence before being allowed to 'pass'. His or her performance relative to other learner drivers should (in principle) be of no consequence.

Norm-referenced assessment

The above approach contrasts sharply with *norm-referenced assessment*, which is altogether more competitive. Norm-referenced assessment involves tests of ability or attainment which are intended to probe differences between individual students, and hence determine the extent to which each individual's performance differs from the performance of others of similar age and background.

In cases where there is a choice of questions in a norm-referenced test, this highlights a need for *standardisation* of scores for comparison purposes. A typical norm-referenced test may have a fixed pass rate (say 55%) which is strictly adhered to no matter how high or low is the general level of attainment. This is, on the face of it, a much less fair approach to assessment than criterion-referenced assessment, since only *relative* attainment, not *absolute* attainment, is recognised. However, the approach is widely used – in many national school examinations and professional examinations, for example.

Comparison of the two approaches

Basically, criterion-referenced assessment and norm-referenced assessment differ in the *purpose* for which the assessment is carried out, the *style* in which the component tests are constructed, and, finally, in the *use* to which the information derived from the results of the assessment is put.

In the remainder of this booklet, we will attempt to demonstrate the role of assessment techniques in a general systems approach to course design. Thus, our main concern will be with criterion-referenced assessment related to the attainment of pre-specified objectives and identifiable behaviours.

Test construction

As mentioned earlier in this booklet, student assessment should be directly geared towards the stated course objectives (while remembering that not all objectives are formally assessable, yet may nevertheless be very important).

The attainment of assessable objectives may be measured in a relatively sporadic programme of *set examinations*, or, more consistently (and possibly less stressfully for students), by some form of *continuous assessment* procedure. However it is done, it is likely that a *combination* of assessment techniques will be necessary in order to assess the range of objectives under investigation validly and comprehensively.

In order to ensure that particular sets of skills are being assessed, some individuals and organisations have drawn up 'tables of specifications' for tests to ensure that due weight is given to all skills and content areas. For example, Figure 2 represents a typical specification of the cognitive skills to be assessed in the UK Ordinary National Certificate (ONC) in Chemistry. The course syllabus is written in the form of behavioural objectives, and the specification is given in terms of Bloom's classification of educational objectives (which is discussed more fully in the booklet on 'Educational objectives') and the various areas of course content. Tables of this sort, while perhaps a little rigid, do enable exam setters to design assessments to cover the full range of skills (in this case, cognitive skills) that are under scrutiny, and to promote good syllabus coverage. They also ensure that certain skills (e.g. factual recall) are not over emphasized, and that due attention is paid to higher cognitive skills.

Subject and Topic \ Ability	1. Recall	2. Comprehension	3. Non-Routine Application	4. Analysis/ Evaluation	5. TOTALS
Inorganic Chemistry					
A. Revision and extension	7	6	4	6	23
B. Chemical Reactions	8	10	2	0	20
C. Group I and II Elements	5	10	2	2	19
D. Group VII Elements	5	10	2	2	19
E. Group V Elements	5	10	2	2	19
TOTALS	30	46	12	12	100
Organic Chemistry					
A. Nomenclature	3	2	0	0	5
B. Stereochemistry	1	6	2	0	9
C. Hydrocarbons	5	7	4	2	18
D. Halogen Derivatives	1	6	3	0	10
E. Hydroxyl Compounds	2	8	3	2	15
F. Carbonyl Compounds	3	10	3	3	19
G. Acids and Derivatives	3	6	2	1	12
H. Bases	2	6	2	2	12
TOTALS	20	51	19	10	100
Physical Chemistry					
A. Gases	4	5	3	1	13
B. Solutions	8	9	6	3	26
C. Thermodynamics	3	4	2	1	10
E. Chemical Equilibrium	3	3	2	2	10
F. Electrochemistry	6	7	4	2	19
G. Ionic Equilibria	6	7	6	3	22
TOTALS	30	35	23	12	100

Figure 2: typical tables of specifications for the UK Ordinary National Certificate (ONC) Examinations in chemistry

The type and range of techniques used within a given assessment strategy will depend upon a number of factors – the most important (at least from an educational point of view) being the student behaviours that are specified in the objectives being tested. The basic characteristics of a range of assessment methods will now be discussed, together with their respective advantages and limitations.

A review of student assessment methods

Student assessment methods can have a wide variety of forms. The most common general approach is via some form of written response, i.e. the *'paper and pencil'* approach. This approach encompasses a whole range of 'traditional' assessment methods such as *essay-type questions*, *short-notes questions* and *problem-solving questions*, all of which require an extended written response of some sort.

Another form of 'paper and pencil' approach involves the use of *'objective tests'*, although such tests seldom involve the student in writing very much; in most cases, a mark made beside one of a range of possible options, or a single word or phrase, is all that is required. Also, the word 'objective', when used in the 'objective test' context, can be somewhat confusing, since it neither means that the questions are necessarily related to the course objectives, nor implies that the questions are objectively chosen. The term simply indicates that the answers to such questions can be marked *totally reliably* by anybody, including non-subject specialists, and, in some cases, even by a computer. The most common type of objective question is the *multiple-choice question* (or, more correctly, *multiple-choice item*), together with its range of variations. Other types of objective questions include *completion items*, *unique-answer questions*, and *structural communication tests*.

Practical tests are often used to assess psychomotor objectives, and include such techniques as *project assessment*, *assessment of laboratory work*, and other *skill tests* designed to assess specific manipulative skills. Also in this category are *situational assessment* techniques, which involve students using non-cognitive skills (such as decision-making skills) in a real, or (more likely) in a simulated environment.

Fourthly, there is a range of *unobtrusive assessment* techniques which can take place without the student necessarily being aware that he or she is in fact being assessed.

Finally, there are the various forms of *self assessment* and *peer assessment*, in which the assessment is carried out by the actual students.

Let us now look at each of these techniques in turn, starting with traditional paper and pencil tests that involve extended writing of some sort.

Traditional 'extended writing' tests

As we have seen, the most common test techniques that fall into this category are *essay-type questions*, *short-notes questions* and *problem-solving questions*. Let us therefore examine these in turn.

Essay-type questions

Essay-type questions are often considered to be one of the 'bluntest instruments' of assessment, having very low reliability and, in many cases, low validity. Often, in a single question, the setter attempts to test *knowledge*, *reasoning*, *written communication skills* (including English language skills, and, perhaps, graphical skills and mathematical skills), *creative thinking* abilities, and *interpretation* (not only of the question itself, but often of the implied objectives of the setter). All of these factors and skills are interwoven in an extremely complicated matrix, and much is left to the judgement (or caprice!) of the marker. Even with the best of intentions, it is almost impossible to tease these skills out and mark them independantly. Even when an *assessment grid* is used, thus enabling the various components of the essay to be marked independently, research has shown that reliability is still very poor, with markers varying widely in their scoring of this kind of question.

Despite this, essays do have a number of points in their favour.

(a) They give students an opportunity to organize their ideas and express them in their own words. Also, scope is provided for the demonstration of written communication skills and for the expression of unconventional and creative thinking. (These opportunities are, however, often lost when 'essays' consist simply of regurgitated class notes).

(b) They allow students to display a detailed knowledge of related aspects of the course being assessed, as well as a knowledge of relevant topics outwith the course proper.

(c) The questions are relatively easy to set.

(d) Many teachers and users of the results of assessments (e.g. employers) hold the opinion that student tests and examinations

should contain at least an element of essay writing (except, perhaps, in mathematical subjects).

Balanced against these advantages, however, are many disadvantages, some of the main ones being the following.

(a) Essay questions are exceeding difficult to mark reliably and, with only one marker, the subjective element can be considerable. The correlation between the scores of two markers for the same set of answers, or even between the scores of the same marker for the same set of answers on different occasions, is seldom sufficiently high to justify confidence.

Essays are also very time-consuming to mark, especially if the marker adds comments and criticisms in order to provide feedback for the student.

(b) Only a small number of long essays can be answered in a given time, thus effectively restricting the assessment to a few (often student-selected) areas of the course content. Other equally-important areas may be completely neglected, and the total mark may therefore be an unreliable index of the student's grasp of the course as a whole. Also, in an examination which consists of a limited number of essays, luck in 'spotting' questions beforehand is often a significant factor.

(c) Where there is a choice of questions, this enables different students to answer, in effect, different papers, so the same total mark may not represent comparable performances. This will almost certainly be the case when the questions vary in difficulty, in content, in the types of skills involved, and are scored by different markers. For example, a '5 from 8' paper contains a total of no less than 56 different combinations in which the 5 questions can be selected!

(d) Occasionally, students may not appreciate the true intent of an essay question because of inadequate direction (e.g. "Write an essay on proteins"). Markers then have the choice of ignoring the answer, accepting the student's interpretation as an answer to a question which was not intended, or adopting an uneasy compromise. Clearly, this adds neither to the reliability nor to the validity of the assessment.

(e) Irrelevant factors often intrude into the assessment, e.g. speed of handwriting (especially with restricted time), style and clarity of handwriting, and grammatical errors.

Detailed guidance on how to write, evaluate and mark essay-type questions is given in a separate booklet ('Essay-type questions').

Short-notes questions

In cases where '*short notes*' on a subject or topic are required rather than an extended essay, many of the problems associated with long essay questions are reduced, although not necessarily eradicated. 'Short notes' questions should (in principle) be more valid and reliable than essay questions, because the marker is able to concentrate more sharply on particular aspects of the answer. In addition, they allow wider coverage of course content, and are generally more specific.

However, although reliability is increased, some deviation in scores may still occur between markers. Also, course coverage may still not be adequate, and students' individual written and presentational skills may again cloud the validity of the questions.

Problem-solving questions

Problem-solving questions are an excellent method of testing some of the middle-to-higher cognitive skills (such as *comprehension*, *application* and *analysis*), and for demonstrating extended reasoning skills. Mathematical, scientific and engineering subjects, in particular, lend themselves readily to assessments of this sort. With such questions, validity may well be high, but problems of reliability may arise in respect of the marking of partially-solved problems.

Objective tests

Objective tests are assessment procedures which can be marked with total reliability. Although such items are often criticized on account of assessing only at low intellectual levels, this is not necessarily the case. It is possible (although more difficult) to design items to test skills in the higher cognitive areas, and even to test logical thinking and skills related to structuring arguments.

Before looking at the characteristics of specific techniques of objective testing in more detail, we will summarize the main advantages and disadvantages of using objective tests in general.

Some of the main points in favour of objective tests are as follows:-

(a) The tests can be marked with complete inter-marker reliability.

(b) Large numbers of questions can be answered, thus ensuring a thorough sampling of course objectives and content.

(c) Objective items can be designed to test specific abilities in a controlled way.

(d) The difficulty of the items is often known from trial-testing. Hence, by selection of appropriate items, the difficulty level of the test can be adjusted to meet particular requirements.

(e) Items can be 'banked' and re-used.

(f) There is no need to provide a choice of questions for the students, and, indeed, this is not desirable, since it tends to reduce validity.

(g) Tests lend themselves to inexpensive and easy marking, and also to thorough statistical analysis. This allows investigation of individual difficulties, and also permits the general problem areas of the student population as a whole to be identified.

Against these advantages, objective items have the following disadvantages:

(a) They are very difficult and initially expensive to construct, and considerable preparation time is necessary. Their *apparent* ease of construction often leads to amateurish attempts, resulting in very poor, invalid items. (This, in turn, has been responsible for some of the criticisms levelled at objective tests). Expert advice is often required in designing items, and all items should be pre-tested in order to measure their level of difficulty and the extent to which they discriminate between the better and poorer students in a given population.

(b) The teacher or marker cannot see the reasoning behind the choice of a wrong answer.

(c) It is difficult or impossible to construct tests to assess certain high-level abilities such as extended reasoning and written communication ability. Thus, objective tests are probably best suited for testing lower cognitive skills, and items at these levels are certainly the easiest to write.

Let us now look at the different types of objective test items that can be used.

Multiple-choice items

Multiple-choice items are probably the most widely used component of objective tests. Several variations on the multiple-choice theme are possible, such as when several items arise out of one situation, graph or set of figures.

The advantages and disadvantages of objective items in general (as listed above) apply in full to multiple-choice items.

An example of a multiple-choice item that is designed to test knowledge is given below:

Which city is the capital of Australia? (mark appropriate box)

(a) Melbourne ☐ (c) Sydney ☐

(b) Brisbane ☐ (d) Canberra ☐

Multiple-choice objective testing has its own associated jargon, the most common terms being as follows:

Stem: the introductory part of the question out of which the alternative answers arise. Ideally, this should be a self-contained question containing all the basic information which the student needs in order to respond to the item, so that he or she does not need to read through the options to discover what is being asked. The stem should be concise, should use unambiguous language appropriate to the students' ability, and should avoid negatives if at all possible.

Options: the range of possible answers. The options should be parallel in content and structure, i.e. they should all have the same kind of relationship to the stem, and should all follow grammatically from it. Obviously, the item should not contain clues in the structure of the options (e.g. mixtures of plurals and singulars).

Key: the correct answer. This must be *unarguably* correct; hence the option 'all of these' should never be used.

Distractors: the wrong answers. These must be unarguably incorrect answers, yet should appear plausible to weaker students.

Non-functioning distractors: those distractors which attract less than 5 per cent of the responses. When an item is re-written, an attempt should be made to replace such distractors with more plausible ones.

Facility value (FV): the fraction (normally expressed as a decimal) of the candidates choosing the key in any given item. Thus, if half the students answer correctly, the facility value for that item is 0.50. In tests of achievement designed to rank students in order of merit, the facility values should lie between 0.35 and 0.85, since *very* difficult or *very* easy items do not normally form effective components of such a test.

Discrimination index: a figure which represents the degree to which the item separates the better students from the poorer students, since a 'good' item (particularly in an achievment test) is one which the better students should get right and the poorer students should get wrong. There are several ways in which the discrimination index

can be calculated, but one of the simplest is to calculate the difference between the facility values for the top third of the population (for the test as a whole) and for the bottom third (again for the test as a whole) for the item under consideration.

The discrimination index can obviously never be greater than + 1.0, and should always be greater than + 0.2 for a 'good' item. A negative discrimination index is a sign of a very poor item that should be either discarded or revised.

When there is a choice of pre-tested items of known quality, the facility values and discrimination indices chosen will depend on whether the test is meant to be of a simple 'pass/fail' type, is meant to produce a meaningful class ranking, is meant to serve as a diagnostic instrument providing feedback on progress for students, or is designed to help evaluate the efficiency of a teaching/learning system.

Detailed guidance on how to write, evaluate and mark multiple-choice items is given in a separate booklet ('Objective questions').

Completion items and unique-answer questions

In both these forms of assessment (which are sometimes collectively known as *short-answer questions*) the testee has to *supply* the answer rather than select from a set of choices provided. Examples are given below:

Completion item 'The United States equivalent of the British House of Commons is known as the . . . . . "

Unique answer question 'What is the equivalent temperature in degrees Celsius to 185° Fahrenheit?'

In both cases, the answer is *unique*, and so the test can be marked reliably; it has, however, to be marked manually. In such items, skills can be examined one at a time, e.g. mastery skills (recall, using formulae, simple calculations, etc), organizing skills (categorizing, etc) and interpretation skills (of graphs, tables, etc). Such items can, in fact, be set at surprising high cognitive levels. Again, relatively full representative coverage of course objectives and content is possible, since only very short written answers are required.

Detailed guidance on how to write, evaluate and mark completion items and unique answer questions is given in a separate booklet('Short-answer questions').

Structural communication tests

Structural communication testing is a development in objective testing in which an attempt is made to carry out a reliable test of a student's ability to select relevant information from irrelevant information and to present structured arguments logically.

Basically, students are presented with a grid containing statements pertaining to a particular topic, all of which are factually correct. The grid can contain any number of statements, but 16 or 20 are typical. Students are then asked questions on the topic, to which only *some* of the statements are pertinent. The student has to select from the grid the *relevant* pieces of information in order to answer the question(s), and then has to *arrange them in a logical order* in order to present the argument. Allowance in the scoring can be made if several logical sequences are permissable. In some cases, structural communication tests can be computer-marked.

Practical tests

Practical tests are highly appropriate in cases when the development of psychomotor or manipulative skills is an important part of a course. Their main drawbacks are that they may be logistically difficult to arrange and administer, and may have low reliability. However, the face validity of actually performing a set task would seem to be high compared (for example) with giving a simple written description of how the task *should* be performed. Let us now examine some of the most important types of practical test.

Project assessment

In such assessment, a student may be assessed in terms of his or her cumulative work over a period of time, or perhaps on only the end result of the project (e.g. a working model, the results of a set of experiments, or a computer program). Such assessment can also be carried out on *groups* of students who have collaborated in carrying out a group project of some sort. As is discussed in 'A guide to the use of group learning techniques', however, this can give rise to problems in assessing the contributions made by the different members of the group unless some form of *peer assessment* is used.

Assessment of laboratory work

In cases where the development of manipulative laboratory skills is an important component of a course (e.g. in science courses), assessment of actual laboratory work may be carried out. This usually takes the form of continuous assessment over a period of time or a one-off practical examination at the end of a course or

section thereof. The latter has the disadvantage that it may be unfair to students who have an 'off-day', and also to students who react badly to exam pressure but have otherwise performed well during the course. From the marker's point of view, it can also be exceeding difficult to monitor the progress of even a small number of students effectively during such an examination.

Skill tests

Tests of the ability to carry out specific manipulative tasks may be important in some courses (e.g. dismantling and reassembling a car engine, cutting hair in a particular way, or repairing a piece of technical equipment). For each of these, a suitable practical test can generally be devised, depending on the circumstances. Such tests are more common in 'training' courses than in general educational courses, however.

Situational assessment

Assessment procedures of this type, which originated in management education, involve the appraisal of complex decision-making skills. They may involve the student in performing such activities as dictating letters, dealing with personnel problems, formulating agendas, and dealing with budgets or financial problems. The situations that are used in such assessment are normally simulated, and a whole range of activities and crises can be 'built in' to arise in the same way as they might in the real world. Such an approach is often called an *'in-tray' exercise*.

Again, the validity of such a technique would appear to be high, but care must be taken in marking the performance in order to ensure reasonable reliability. To this end, a checklist containing the objectives under assessment provides a useful guide for the marker.

Unobtrusive assessment

Unobtrusive techniques involve the students being observed and assessed without their prior knowledge. Such techniques can be important in assessing a student's *commitment* and *attitudes* to work, rather than simply his or her ability to perform tasks under the controlled conditions of more formal assessment. They can, therefore, be more valid than (for example) written examinations, which invariably have a large element of artificiality. On some occasions, video techniques are used for recording student performance and for subsequent analysis and assessment of personal skills and traits. However, there are often considerable logistical problems in operating such an approach, not to mention the obvious doubts over the ethics of unobtrusive assessment.

Self and peer assessment

The idea of allowing students to assess both their own work and the work of other students is currently gaining ground in higher education. One argument in favour of self-assessment is that we should be encouraging students to become more self-critical and more able to judge the worth of their own work. After all, it is likely they will be expected to do this is later work situations. Experience in the use of self-assessment has resulted in the (perhaps expected) finding that students generally do not overmark themselves compared to tutor marking. Indeed, the correlation between the two is very good, and, if anything, students tend to mark themselves *downwards*, and are often extremely critical of their own work. Obviously, some preparation is required before such a scheme can be adopted, involving, among other things, negotiation between tutors and students regarding the criteria for assessment and their relative weighting.

Peer assessment is mainly used in group-based projects or other collaborative exercises, when students may mark one another in respect of their individual contributions to the combined work. Again, negotiation of criteria is necessary, and there is again the possibility of mutual overmarking, although experience so far indicates this is not an overriding problem. Indeed, it can be argued that the benefits of increased motivation and self-awareness greatly outweigh any such possible disadvantages.

Conclusion

If one has an area of course content, a list of objectives and an exam specification of required skills, it should be possible to construct a valid programme of assessment by selecting those objectives which can be tested by objective items, those which require short written notes, those which require to be assessed in a practical setting, and those which lie in the area of attitudes and disposition. The few objectives left over (e.g. those involving extended reasoning or written communication skills) may then need essay-type questions. If everything else has been dealt with by more appropriate methods, the marker can concentrate comparatively single-mindedly on these few areas in the essays, thus tending to make marking more reliable.

In short, an appropriate *battery* of assessment techniques should be used to match specific objectives, thus producing a *practicable* assessment strategy that not only has a high degree of *validity* and *reliability*, but is also *fair and useful* to students.

Further Reading

1. *Essentials of Educational Measurement*, by R L Ebel; Prentice Hall, Englewood Cliffs, New Jersey; 1972. (One of the definitive texts on assessment; an extremely useful guide to the field.)
2. *Assessment Techniques*, by B Hudson; Methuen, London; 1973. (Another extremely useful basic text.)
3. *Assessing Students – How Shall We Know Them?*, by D Rowntree; Harper and Row, London; 1982. (Another extremely useful book that deals with all the material covered in this booklet in much greater detail; also highly recommended.)